



University of Pennsylvania
ScholarlyCommons

GSE Publications

Graduate School of Education

2016

In Search of *Terra Firma*: Administrative Records on Teachers' Positional Instability across Subjects, Grades, and Schools and the Implications for Deploying Randomized Controlled Trials

Robert Boruch

University of Pennsylvania, robertb@gse.upenn.edu

F. Joseph Merlino

Jill Bowdon


University of Pennsylvania

John Baker

Jessica Chao

See next page for additional authors

Follow this and additional works at: http://repository.upenn.edu/gse_pubs

 Part of the [Curriculum and Instruction Commons](#), [Educational Administration and Supervision Commons](#), [Educational Methods Commons](#), [Education Policy Commons](#), and the [Teacher Education and Professional Development Commons](#)

Recommended Citation

Boruch, R., Merlino, F., Bowdon, J., Baker, J., Chao, J., Park, J., Frisone, M., Ye, T., Hooks, T., & Porter, A. C. (2016). In Search of *Terra Firma*: Administrative Records on Teachers' Positional Instability across Subjects, Grades, and Schools and the Implications for Deploying Randomized Controlled Trials. Retrieved from http://repository.upenn.edu/gse_pubs/393

This is an Original Manuscript of a forthcoming article to be published by Taylor & Francis in the Journal of Research on Educational Effectiveness.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/gse_pubs/393

For more information, please contact repository@pobox.upenn.edu.

In Search of *Terra Firma*: Administrative Records on Teachers' Positional Instability across Subjects, Grades, and Schools and the Implications for Deploying Randomized Controlled Trials

Disciplines

Curriculum and Instruction | Education | Educational Administration and Supervision | Educational Methods | Education Policy | Teacher Education and Professional Development

Comments

This is an Original Manuscript of a forthcoming article to be published by Taylor & Francis in the Journal of Research on Educational Effectiveness.

Author(s)

Robert Boruch, F. Joseph Merlino, Jill Bowdon, John Baker, Jessica Chao, Ji Eun Park, Michael Frisone, Tianpeng Ye, Tom Hooks, and Andrew C. Porter

In Search of *Terra Firma*: Administrative Records on Teachers' Positional Instability across Subjects, Grades, and Schools and the Implications for Deploying Randomized Controlled Trials¹

Robert Boruch, F. Joseph Merlino, Jill Bowdon, John Baker, Jessica Chao, Ji Eun Park, Michael Frisone, Tianpeng Ye, Tom Hooks, Andrew Porter

5/2/2016 5:30PM

In their dogged quest for structural integrity, engineers don't build in swamps unless they know how to account for unstable ground. Often, however, education researchers do the equivalent of trying to build on swampy land. In particular, we design randomized controlled trials to test school based interventions in ways that often fail to adequately address teacher movement, or instability, across subjects, grades, and schools. Such teacher shifts, referred to here as *ambient positional instability* (API), poses a number of challenges for educational researchers.

API can weaken the foundation of a trial, for instance, when the intervention depends on teacher stability. Such stability is crucial in interventions where teachers receive professional development in a particular subject at a certain grade level with the goal of improving student achievement. In controlled trials on the effects of interventions that are otherwise well designed, this instability can reduce the size of the intervention's apparent effect and the chance that the effects will be discerned. Beyond this immediate concern, of course, API may affect children's achievement in a trial and in observational studies of achievement.

¹ The research on which this report is based was supported by Grant DRL-1337237 from the National Science Foundation. The views expressed here are those of the authors and not necessarily those of the Foundation

This article reports on early findings from a study of API as it pertains to randomized controlled trials. The current lay of the land is mapped out first, considering illustrative studies that arguably depend on teacher stability. We then present findings briefly on an analysis of articles published in the *Journal of Research on Educational Effectiveness* (JREE), noting how API is reported and not reported in the context of experiments.

Most important, new work on API is reported, based on administrative records on entire populations of teachers from six states and their big cities. These constitute “bigish,” though not big data, and that are complex. The results include magnitudes of API among teachers over 5-16 year periods depending on the state. Information on the accessibility of relevant records is also provided. The closing remarks consider the implications of the findings for designing and conducting randomized controlled trials, and for statistical policy more broadly.

Ambient Positional Instability: A Refined Approach

People who are involved in public education, including researchers, often take teacher movement, within and between schools and among subjects and grades, as a fact of life. Though the disappearance of teachers from a controlled trial is often labeled as “attrition from the study,” this designation is crude. Here, we use the phrase ambient positional instability (API) to enlarge the boundaries of what it means for teachers to disappear from trials. Our conception of teachers disappearing from a study includes conventional indicators of teacher movement between schools and districts and their departure from the profession.

We go beyond convention to include teacher assignment changes within schools and across schools in the subjects they teach, as well as changes in the grade level they teach over time. Further, we get beyond conventional dependence on sample surveys of teachers and teacher samples in local

trials and focus on entire populations of teachers over time. Being attentive to subject and grade assignment changes is important because many randomized controlled trials involve interventions that are subject and/or grade specific.² Public records on populations of teachers and computed API rates are a potentially valuable resource in designing trials and interventions. Though API can entail changes within-year, we focus here on changes across years. When aggregated, these various types of shifts constitute what we mean by API.

Randomized Trials and Teachers' Ambient Positional Instability

In designing multi-year interventions and the randomized trials that test their efficacy, teacher instability can have serious implications. Consider, for instance, the influence of API on a large study conducted from 2008 to 2014 by the National Research and Development Center for Cognition and Science Instruction (Scull et al 2015; Yang et al 2015). The arrangement of cluster-randomized trials involved 180 schools and tested whether cognitive science principles used to modify existing middle grades science curricula had effects on student achievement. Such modifications were expected to lead to improved student learning. Two modified curricula were implemented in two grades in 60 intervention schools. Another 60 schools received an alternate treatment of professional development designed to improve teacher science content knowledge. A third set of 60 schools served as a business-as-usual control. More than 540 middle grade science teachers were initially engaged. The test of the intervention was to take place in the *second* year of implementation, after teachers had received a full year of professional development (summer and academic year) and had a chance to experience and practice implementing the cognitive science interventions during the first year.

² API at higher levels, such as for principals and superintendents, is also important. So is student mobility. See, for instance, Hanushek, Kain, and Rivkin (2004) on administrator changes and Mehana and see Reynolds (2004) and Vuchinich et al. (2012) on student mobility. In this article, we confine our attention to public school teachers.

Conventional measures of teacher turnover or attrition recognize only teachers leaving their school. For this trial and others, however, a science teacher who switched teaching assignments from science to math, or from eighth grade to fourth grade, had effectively left the study, just as if they had left the school. A large northeastern school district with 92 middle grade schools participated in one of the two trials. Nearly 42% of its science teachers who participated in the first year of the intervention beginning in the summer of 2009 had switched grade, subject, or school by the start of the second year (September, 2010) —the year the intervention was supposed to have its maximum effect. As one of the researchers remarked, “Trying to provide effective professional development to these middle school science teachers is like trying to golf in a hurricane.”

The idea that teacher instability is a silent but crucial factor in an intervention’s potential effectiveness is evident in other evaluations of multi-year educational interventions. For instance, in Garet et al.’s (2011) randomized trial of a professional development intervention in middle school math, roughly half the teachers who were recruited into the intervention and control conditions in the first year left the activity by the end of the second year. Springer et al. (2010) reported results of a randomized trial in Nashville that was designed to determine whether bonuses paid to teachers would result in higher achievement of children in mathematics. Of the nearly 300 teachers initially engaged, half left the study by the end of the third year. There was no difference in rates of leaving between control and intervention/bonus arms of the trial.

A trial by Hanson et al. (2012) involved 50 schools in a study of the effect of a character education program on fifth-grade students’ achievement. Between Year 1, in which random assignment was done, and Year 2 of implementation, 18% of teachers who began in the study had moved out of the school, were assigned to an ineligible grade, or withdrew from the study. Moreover, about 23% of the teachers who had joined for Year 2 of the study moved out of the study’s ambit by the end of the year.

Finally, consider Heller's (2012) report on a randomized trial of a professional program for middle school science teachers in California and Arizona. Of the 181 teachers who were initially engaged and randomly assigned to the Making Sense of SCIENCE program and to control conditions, 48 (27%) "left the study before data collection was completed" (p. xi) between 2009 and 2010. Here again, there was no appreciable difference in the rates of departure across the arms of the trial.

API appears to be not much of an issue in randomized trials lasting less than a year. Weijekumar, Hitchcock, Turner, Lei, and Peck's (2009) cluster randomized trial of the Odyssey Mathematics program for fourth graders in a sample of schools in the mid-Atlantic region is a case in point. None of the 122 teachers in 32 schools disappeared from the trial over its course. One plausible reason for this is that the trial was brief in duration. A second possible explanation might be that the sample of schools involved was suburban rather than urban. More about this anon.

Characterizing instability in the local system is of course not the only vulnerability in tests of interventions. The topic fits into a larger context of research methods for anticipating failure and understanding it well. See, for instance, Boruch and Ruby's (2015).

Lacunae in Reports on Randomized Trials

Reports on randomized trials that provide information on API indicators, such as those cited above, are exceptional. As part of our current project, we reviewed all articles published over 5 years in the *Journal of Research on Educational Effectiveness (JREE)* to learn whether teacher instability is addressed at all and, if so, how it is handled. Over 30 articles provided results of randomized controlled trials. Most of these involved a one year intervention. Only 10% percent of the reports attended to API at all, regardless of the trial's results. *Post facto* explanations for failure to detect effects were at times put forward in articles when this was the result of the trial. The reasons offered for finding no discernable effects included weak implementation, poor fidelity, teachers' experiences, or irrelevance of

the outcome measure. These speculations are important. But they seem insufficient unless instability in the teaching cadre is recognized.

There are several plausible reasons why researchers may not report on API in their trials. First among them is that many trials involve short time frames. Figure 1, based on Park and Boruch (2014), shows the duration of interventions covered in reports on randomized trials that appear in *JREE*. In particular, among the journal articles we reviewed, the majority had a time frame of a single school year or less. API from one year to the next is then irrelevant for such interventions. Although within-year instability is relevant, this was reported in only a few articles.

A second explanation for why authors do not consider API in the context of controlled trials may be their presumptions. A principal investigator from a well-regarded research firm, for instance, said during an interview, “I just consider it part of what I have to deal with in the study; it is a given.” The investigator did not regard API as a phenomenon worthy of research, of a methodological sort or otherwise.

A third reason that API might not be considered in reports is that contemporary reporting standards, though excellent in many respects, do not require researchers to account for it. For instance, the Institute of Education Sciences’ (IES) standards of evidence for the What Works Clearinghouse (WWC) attend well to attrition/retention of students in the trial. But these standards do not attend to other aspects of system instability, such as shifts in teachers’ assignments to schools or to academic subjects. Nor do they attend to potential implications of instability for effect size and for judging external validity (generalizability) of findings, among other issues.

Other researchers, of course, have considered aspects of teacher API. A few of the major resources are identified next and contrasted with the approach to API taken here.

Reports of Educational System Instability More Generally

The National Center for Education Statistics (NCES) Schools and Staffing Survey (SASS) and the associated 1-year Teacher Follow-up Survey (TFS) are important resources in understanding mobility among a national sample of teachers. The results are generalizable to the nation and applicable to particular time frames. They are not generalizable down to the level of State Education Agencies (SEAs) and Local Education Agencies (LEAs), however. Consequently, they are not directly useful for determining in which districts or individual schools there is enough teacher stability to conduct a multi-year randomized controlled trial effectively.

Some studies of conventional turnover do focus on particular districts, but they tend to be episodic, unlike the regular 3-year cycle of the SASS and TFS. For instance, a Research for Action report by Useem, Offenber, and Farley (2007) on Philadelphia schools tells us that of *new* teachers hired in 1999-2000, only 16% were in the same school in 2005. Of teachers hired in 2004-2005, only 68% were in the same school by the end of the second year of their tour. Tobias (2012) collected data on New York University graduates hired to work in New York City schools in 2006 and 2007. He reported that estimates of rates of shifts from their initial positions were about 27% after 1 year and 38% after 2 years in all schools and grades K-12. Marinelli's (2011) report for the Research Alliance for New York City schools gives data on all teachers in all schools from 2002 to 2009. Middle school teachers departed at a higher rate than others, with about 45% of them leaving their schools within 2 years of entry, and 57% departing within 3 years. The range of rates of departure within years across middle schools is remarkable, from a high of 66% in Manhattan to a low of 35% in Staten Island, as measured against a baseline year.

These studies are informative, but none of these studies report on shifts of teachers from one grade to another or from one subject assignment to another, which is usually the target of an

intervention. Nor do they depend on population data as opposed to probability samples (national level) or idiosyncratic ones (as in local trials). Getting beyond these limits is the next topic.

Public Resources on Population Records on Teacher API and Illustrative Findings

The focus in what follows is on publicly accessible multi-year records on entire *populations* of full-time public school teachers at the State Education Agency (SEA) and Local Education Agency (LEA) levels. This emphasis, of course, is different from probability sample surveys carried out at the national level. The focus is based on the idea that it is at the LEA level that controlled trials are mounted. It is at the SEA level that some local trials are put into regional contexts. Moreover, it is at the SEA level that public records of teacher shifts in assignments from one grade to another, or one academic subject to another, as well as from one school to another, are sometimes maintained and can sometimes be acquired. Such information is, of course, important for other purposes, including research on human capital, studies of school-reform effects, and studies of the relation between API and student achievement (Bowdon, 2015; Bowdon and Boruch, 2014; Merlino and Musa, 2014).

The following sections provide illustrative analyses of population records acquired at the state and big-city levels for Missouri, Ohio, Illinois, Pennsylvania, Arkansas, and New Jersey.

API in Missouri and Its Big Cities

Consider, first, a teacher cohort retention rate based on administrative records from Missouri's five biggest cities: St Louis, Kansas City, Springfield, Independence, and Columbia. Understanding the magnitude of instability and how this varies across different types of schools can help researchers appropriately design multi-year experiments and recruit enough participants to sustain sufficient statistical power. For example, researchers designing a multi-year study focusing on professional development for middle and high school STEM teachers in the largest cities in Missouri should be aware

that one-quarter or more of the teachers in a given year no longer teach the same subject by the following year (Shown in Figure 2 and based on Bowdon and Boruch's (2014) report, ³).

For instance, of the 581 math teachers in the five biggest cities in Missouri in 2005, only 72% were still teaching math in 2006. And of the 537 science teachers in these cities in 2005, only 74% were still teaching science in 2006. While the proportion of science teachers in the largest cities who were still teaching science in the following year remained fairly stable over time, this was not the case for math teachers. In 2012 as few as 60% of the 641 math teachers in the largest cities in 2011 were still teaching math.

If a researcher wanted to design a study to take place throughout the state over several years, it would be better to use different indicators of instability, notably a cohort retention rate at this level, for different types of schools in order to produce reasonable predictions of the necessary sample size.

Figure 3 portrays the cohort retention rate over a ten year period. It shows that, after one year, 70% of STEM teachers were still teaching STEM in the same school in the largest urban districts, while 80% of STEM teachers in the rest of the state were still teaching the same subject in the same school. By 2009, there was a 20 percentage point difference between attrition in the largest urban districts and in the rest of the state: only 40% of STEM teachers were still teaching STEM in the same school in the five largest urban districts, while 60% of STEM teachers in the rest of the state were still teaching STEM in the same school. Attrition, the complement of retention, was higher in urban areas than in the rest of the state, and this disparity widened over time.

API in Ohio and Its Big Cities

Ohio's public web-accessible records can, at times, be used to examine year-to-year stability among all teachers in the state and cohort retention. Baker and Boruch (2015) acquired and examined

³ These data were supplied by the State's School Data Director, Tom Ogle, in the Office of School Systems Data Management.

records on all of Ohio's science and math high school teachers over a 5-year period.⁴ The findings relate to the population of all the state's teachers who were employed in their positions in the 2008-2009 school year. The numerical identifiers for teachers in the files are unique and consistent from one year to the next. Table 1, excerpted from Baker and Boruch (2015) summarizes statistical results for this population.

In the baseline 2008-2009 academic year, of the 1,465 math and science teachers who were teaching at non-charter public schools in the state's five biggest cities, 25% of them had left their schools to do something else in the following year (2009-2010). The "something else" included teaching different subjects or grades, teaching at a different school or leaving the system. Two years later, in 2010-2011, 45% of the initial cohort had left their initially assigned school. In that same year, only half the teachers were teaching the same subjects in the same schools compared to the baseline. Five years later, in 2013-2014, only 25% of the initial cohort of teachers were teaching the same subjects in the same schools.

As is the case for Missouri, the teacher stability rates in public high schools for the biggest cities in Ohio are in sharp contrast to the stability rate in analogous schools apart from others in Ohio. At the five-year point, for instance, the cohort retention rate is 50% (twice as high) for the 9,888 teachers in non-charter high schools outside the cities teaching the same academic subjects in the same schools as compared to the positional retention in the cities. That is, this indicator of API rate in Ohio's biggest cities is roughly twice the rate of API outside the cities.

Ohio's public records allowed the calculation of teacher stability rates of the State's charter high school teachers, a subpopulation of considerable interest to some educators. The positional retention rate over five years of the 625 math and science teachers at charter high schools was 15 percentage

⁴ Ohio records are accessible at <http://reportcard.education.ohio.gov/Pages/Power-User-Reports.aspx>.

points lower than the rate for non-charter math and science teachers in the cities. In Ohio then, this subpopulation of teachers is the most susceptible to positional instability.

The publicly accessible data files on Ohio teachers provide only very general entries on the areas in which they teach, e.g. math or science. Given the broad characterization, it is impossible to track teacher's shifts across specific course assignments. Consequently, we focused on teachers who taught math only, science only or both, so as to understand rough transitions. This revealed that after five years, 4% of the public non-charter high school teachers who remained in their initial school had shifted their broad assignments. The percentage goes up to 11% for high school teachers in the five biggest cities and up to 13% for the charter high school teachers. A finer grained teacher assignment variable has the potential to uncover high rates of shifts across courses within schools.

API in Illinois and Chicago

Consider next Illinois' public records on teacher populations in Chicago and statewide from 2007-2008 through 2011-12. Matching teachers across years and schools etc. was based on teachers' full names and other information, inasmuch as the records do not contain unique numerical identifiers. Table 2, adapted from Chao, Park, and Boruch (2014), portrays the data for teachers in elementary, middle, and high schools.

Illinois' data system for the period include records on just under 107,000 teachers. Of the nearly 20,000 full-time teachers who taught in any Chicago school in 2007-08, only about 8,700 remained in the same school five years later, a loss rate of 57%. Chicago's API rate of 57% is then higher than the API for Illinois as a whole (38.6%): Chicago's API is 1.4 times that of the state. The rate is also higher than the API rate for the state's five largest school districts outside of Chicago (40.3%).

Figure 4 portrays the instability for four academic subject assignments statewide between 2007-2008 and 2011-2012 for the high school teachers in the 2007-2008 cohort. It is only for the high schools

that one can learn about teachers' specific assignments to teach particular subjects based on the code books for Illinois public data files. The population base for the cohort's math and science high school teachers was just over 8,000 teachers. The patterns of teacher instability across subject areas in the cohort differ in that Science teachers were most stable (71.8% retention) and English teachers the least stable (61.3%). Math and social science rates are about the same at 66%.⁵

Technical issues in analyzing these records are given in Park, et al. (2015) and Chao et al (2014). The issues include matching teachers' records across years, inasmuch as the State's public records contain only teacher's names, rather than unique and consistently used alpha-numeric identifiers. Name changes and duplicate names, of course, complicate the task of matching records.

API in Pennsylvania and Philadelphia

Some of the API-related rates for the 2008-2009 cohort of math and science teachers in Philadelphia's non-charter secondary schools are graphed in Figure 5. Weathers and Baker (2015) and their colleagues acquired the public records and presented results at a research-to-practice conference for the Philadelphia School District. In this cohort, there were 1,896 teachers in the population records to which we had access. Over a five-year period, roughly 80% of these teachers shift schools or teaching assignments, or both, or leave for other reasons. The percentage of the 2008-09 teacher cohort who taught the same subject in the same schools drops in each of the four subsequent years: 46%, 33%, 22%, 19%. In other words, over a two-year randomized trial, one may lose 54% of the teacher who were initially assigned to teach a specific subject in their school.

As in other states, the high instability rate in the Pennsylvania's urban schools is in contrast to the state-wide rate. Figure 6, for instance, illustrates a state-wide retention rate of 51% over a five year

⁵ Chao led the search for relevant public records on this. Nonetheless, we have benefitted from the Chicago Consortium for School Research whose products have informed our own work.

period in teachers' staying at a particular school with the same math or science assignment. This retention rate is based on the records of the 17,491 relevant teachers in the State's public record system.

Higher instability rates occur in Philadelphia and are not just restricted to specific teaching assignments. Each section of each bar in Figures 5 and 6 tend to be wider for Philadelphia teachers than for all teachers in the state with the exception of being the section of the bar indicting teachers who retained the same assignment in the same school. This suggests that Philadelphia teachers are more likely to change assignments within a school, change schools within the Philadelphia district or leave teaching in Pennsylvania altogether relative to all teachers in the state. The API rates for math do not differ appreciably from the API rates for science in Philadelphia, nor do the API rates for teachers who teach these two subjects differ in the state as a whole. But the differences between the city and the state API for these subjects are remarkable. Figure 7 is not as fine-grained as the earlier ones, but it does dramatize that difference.

API In Arkansas

Records on the population of all public school teachers in Arkansas for the period 2010-11 through 2015 were acquired in 2015 following negotiations with senior administrators in the State's Department of Education. Details on acquisition, data screening and cleaning, corroboration of exploratory analyses, and deeper analyses are given in Frisone et al (2016).

Table 3 below gives the rates of stability for the 2010-2011 cohort of high school teachers. The table categorizes only high school teachers who are identified in the records as teaching one of the core subjects of interest here: Math, Science, English, and Social Studies. The particular focus is on teachers who remained teaching the same core subject in the same school over a five year period. From Table 3, one learns that that about 40% of teachers in the 2010-2011 cohort remain in the same subject area and the same school five years on, in 2014-2015. Cumulatively, 60% go on to do different things over the

period. Further, there is no appreciable difference in the rates of retention across the subject areas taught for these high school teachers.

Table 4 provides evidence on the year to year churn (instability) among the population of all teachers in all schools in Arkansas for the same time period. From one year to the next, 30-40 % of all Arkansas teachers shift in the subjects that they teach. The rate increased from 30% to nearly 40% over the period. Frisone et al (2016) give data that reflects the fact that the increase in churn has been driven more by new hires than by teachers leaving their assigned positions.

Frisone et al (2016) focused on churn in the sense of teachers moving into or out of a school (newcomers and leavers). From year to year, the rates at the state level range from 37% in the earliest years to 47% most recent year. Little Rock, Arkansas' biggest city, has the highest churn rate, moving from 44% in 2010-2011 to 65% in 2014-2015. The churn rates at the school level for middle schools and secondary school teachers is over 1.3 times the churn for elementary school teachers.

API in New Jersey

Public records on the population of full time teachers New Jersey were acquired for the period 1996-1997 through 2011-2012. For this period, and over time, the number of uniquely identifiable public school teachers increased each year from about 95,000 to over 125,000. This resulted in examination of records on about two million teachers. Because unique numerical identifiers across time for each teacher in New Jersey were not accessible at the time of file retrieval in 2015, unique identifiers were constructed from on record contents that included people's names and auxiliary information such as date of birth. The following information is based on presentations to the NJ Department of Education (Ye et al, 2016 a; 2016 b) and other reports. We are grateful to colleagues in the New Jersey Department of Education, notably Michael Keith, James Riddlesberger and Shannon Tootell for facilitating access to record in the data system on each.

Of the 1996-1997 teacher cohort's 60% remained in the same school for five years. Only 14% remained in the same school for 16 years. Though the profile of reduction in retentions is similar for the big cities and the state, Newark and Camden's school level retention rates are lower than other cities and lower than the state rates (Figure 8, Ye et al. 2016a,2016b).

The churn rate from year to year among New Jersey teachers of core subjects ranges from 20% in early years (1996-1997) up to nearly 40% in 2008-2009, and drops to 25-30% afterwards. The rates for math, science, English, and social studies do not differ in magnitude or profile over the 16 year period of study. These rates for core subjects is greater than the churn in the state as a whole at 20-25% (Figure 9, Ye et al. 2016a, 2016b).

Implications

Consider next the implications of this research for designing and executing randomized trials on education innovations. This section poses a series of questions and tentative answers that have emerged from analyses thus far. A few implications for education statistics policy are then deduced.

Implications for Randomized Controlled Trials

Should experimenters attempt to obtain data on stability of teachers in the proposed schools involved in a randomized trial prior to launching the study? Our reasons for answering this question in the affirmative are straightforward. If one assumes that teachers will be around for a year or two after their being provided with professional development on the new program of interest, and will then engage in the program (and control conditions of course) in the same grade and subject/course, and if they do not stay around, the trial will be challenging in execution and analysis. Acquiring records in advance on teacher instability can avert problems. From conducting microanalyses of API at particular schools, it has become clear that certain schools may be too risky due to their instability to conduct randomized controlled trials, at least not without serious consideration of what might ensue.

Are data on instability available? We would answer with a qualified Yes. We have been gathering data for almost two years and are still acquiring public records from other states. Learning whether one can indeed acquire dependable evidence on the instability ahead of a trial is itself a challenge. One might construe instability among a teacher cadre as a kind of non-compliance in school-based cluster trials. An implication of this is that cluster-level variables that correlate with instability, such as poverty level, might be observed and used to impute data at the school level or at teacher level for teachers who disappear from the trial. This idea, suggested by Vuchinich et al (2012), seems not yet explored.

Should researchers try to do controlled trials only in education systems that are stable, based on the population records for the jurisdiction that is targeted? If the answer is Yes, this has further implications for external validity. Obviously, evidence that some interventions work in stable environments is useful. But the same interventions may not work in unstable ones. The further implication is that *prior* to a full-blown randomized trial, reconnaissance is essential if only to understand the API in the sites targeted for a trial. As a matter of research policy in the United States, at least, the sponsors have to understand that this pre-experiment reconnaissance requires resources. Certainly, understanding how to incorporate related information into statistical power analyses in designing trials is essential.

Should randomized trials be limited to testing interventions with duration of a year or less, rather than interventions that must be deployed over 2 or 3 years? If the answer is Yes, the across-year instability problem disappears. Of course, if one believes that education interventions must be introduced by the same people over longer periods of time to be effective, then other options have to be considered.

One could, of course, try to dodge dependence on teacher stability by inventing and testing other options. For instance, peer-assisted learning and parent education can be construed as ways to

get around the instability of teachers and other “moving parts” of a formal school system. Such programs have met with some success. See, for example, reviews of controlled trials in this area produced by the international Campbell Collaboration at <http://campbellcollaboration.org>. Of course, parent and student mobility will then also be a challenge.

Can we build theories of effect in field trials that can then be corrected for API? Consider, for instance, a simple arithmetic simulation of how the effect size uncovered in a local trial may be degraded in moving up to a sequence of larger field trials, i.e. in a scale-up. Suppose that the intervention in the local setting works in moving children from the 50th percentile to the 65th percentile in their achievement scores over one year. This can result in a happy and justified declaration that the effect size for the intervention is 1 standard deviation and, assuming a decent sample size, is statistically significant.

A simple simulation posits that an effect size found in this local trial will be reduced by the API rate in each subsequent effort to scale up. A locally discernable effect of 1.00 may be reduced by 20% in a new field test that runs over two years in which the API = .20. The initial effect size might be reduced still further when tested in a larger, multiyear trial in which the API is 50%, reducing the detectable effect size. And if API among students and principals is taken into account with the same arithmetic, experimentalists will have to detect effect sizes of .05 or less.

However, what happens to this effect size when an effort is made to scale up? Our simple simulation posits that an effect size found in a local trial will be reduced by the API in each subsequent effort to scale up. Reckoning whether this small theory holds up depends on organizations that accumulate evidence from multiple trials that are run over multiple years on the same intervention. These organizations include the international Campbell Collaboration, the What Works Clearing House, the Coalition for Evidence-Based Policy, Slavin’s Best Practice summaries, and others. None are currently set up to produce data that could easily be used to test any theory of API, but they have

promise. In particular, reviews of studies could routinely include attention to indicators of API reported in each so as to better map the terrain.

Should experimenters behave like engineers, who are routinely confronted by unstable conditions and develop better ways to build their studies to withstand instability? When teachers disappear from each arm of a controlled trial, for instance, they must then be replaced by new teachers who themselves must be trained. New teachers involved in an intervention, for instance, may have many things to do—finding the restrooms, handling hallway mayhem, and coping with acting out students. This is in addition to figuring out what he or she is supposed to do to implement the new program being tested.

Determining which aspects of API may be controllable in a given context can be construed as a part of such engineering. Can incentives, such as “paying for staying,” or administrative remedies, such as induction and retention programs, reduce API? This, of course, requires some serious thinking about what level of API is desirable, quite apart from organizational control devices. We have at least anecdotal evidence that local administrators can be made to understand what incentives matter for which teachers and how resources can be used to actualize the incentives. Fullan and Hargreaves (2012) put the issue in more general terms of developing “professional capital,” including incentives such as ensuring economic returns in the teaching profession, and developing a culture/profession of commitment and the capacity to make effective judgments.

Are there implications for generalizability and replication of results of a set of randomized trials? The idea of generalizability typically involves determining important characteristics of the contexts (sites) in which experiments have been embedded, and then locating other contexts that have similar characteristics and to which results of a set of trials might be generalized. The process may employ sophisticated statistical methods for matching sites, including the use of propensity scoring methods to identify probabilistic matches, e.g. Stuart et al (2011), Tipton (2014), and Chan (2016).

It is reasonable to consider indicators of ambient instability, such as teacher retention and churn in the schools in which cluster trials are run, as matching variables in any attempt to generalize to other schools. Tipton's (2014) work has been used to develop software, for instance, that assists; see <http://thegeneralizer.org>. Similarly, concerns about replicability and replication of a cluster trial in schools may also benefit from considering indicators of instability. Anticipating failure to replicate or explaining such a failure post facto for instance may depend on these. Further work on this is underway.

Implications: Federal Statistical Policy

Many readers understand the high value of the Common Core of Data (CCD), the Schools and Staffing Survey (SASS), and the Teacher Follow-up Survey (TFS). The latter is, for instance, important for characterizing conventional teacher turnover (leavers and stayers) based on well-designed national probability samples. But estimates at the national level are of no real use at the SEA and LEA levels, where the aim is to test innovations deployed in multiple schools in multiple cities and directed at particular academic subjects or grades. This prompts another question:

Can administrative records on the entire population of teachers, generated and maintained by states, be used to augment, supplement, or substitute for parts of contemporary probability sample surveys? How might they lower the cost of such surveys?

Such records are not uniform, nor are they easily accessible or understood. Each State's data system presents challenges in terms of physical access, as well as understanding coding schemes, definitions, organization, rules of engagement, and other factors that influence one's use of the records. This matter, which is endemic to any effort to share information, invites our final question:

How might statistical policy at federal or state levels be configured so as to foster ease of access to and interpretability of public records? We leave this to our colleagues in the federal government to

address this query, in cahoots with the State people and organizations that produce or use the information.

Concluding Remarks

The first objective of the research reported here was to determine whether public records on ambient positional instability of public school teachers could be acquired. The second was to produce informative descriptive statistics at the population level for schools, districts and states, if indeed the records could be acquired, and to deduce some implications of acquisition and analysis.

We have found that the records on the entire population of teacher position assignments over substantial periods of time can indeed be acquired for the States that we targeted. The capacity to acquire records constitutes a major advantage in the context of controlled trials. Because they pertain to school and district levels, the records can inform decisions about whether, when and where to undertake multi-site, multi-year trials. They can enhance one's understanding of the problems one may encounter in the trial's execution and in analyzing resultant data.

To the extent that designing school-based interventions depend on teacher's positional stability, the particular design can be informed by relevant population records at the schools level. Indeed, if designs can be structured so as to take instability into account or manage it, getting hold of these types of assignment records and analyzing them for teacher stability may not only be helpful, but essential.

Beyond the obvious, the use of public records on teacher assignments can elevate the credibility of attempts to estimate the effects of interventions based on non-randomized (observational) studies. Acquiring and making use of these public records entails a relatively low cost, relative to other expenses of conducting a field experiment. It is especially important at the jurisdiction levels (school or district) where poverty, among other factors, can be related to teacher instability and, in turn, to student achievement.

If public records could be linked, at times, to relevant federal surveys of the kind run by NCES or to local surveys run by others, the cost of running such surveys might be reduced, and insofar as a randomized trial engenders two parallel surveys at times, the cost in these might also be reduced. Attending to ambient positional instability is important on common sense grounds and on account of evidence and theory at hand. If experimenters fail to exploit the public good – records on teachers – we will fail also to do an excellent job in designing randomized controlled trials on STEM Programs and in other areas.

Acknowledgement

The work reported here was funded by National Science Foundation Grant DRL-1337237. We are grateful for comments from Brian Flay, Jennifer Moore, and Renya Wasson. The views expressed here are those of the authors. Nobody else is blameworthy.

References

- ACTION United Education Fund. (2012). *Revolving doors: Findings from Philadelphia's highest poverty schools. Part 1: principal turnover and the importance of stable leadership*. Philadelphia, PA: Action United. Retrieved from <http://www.actionunited.org>
- Baker, J. Y., and Boruch, R. (2015). Ambient Positional Instability among Ohio Math and Science Teachers: 2008-214. Working Paper/Briefing API -Ohio. Conshohocken PA: 21PSTEM and Philadelphia: University of Pennsylvania. Accessible at: http://repository.upenn.edu/gse_pubs/269
- Boe, E. E., Cook, L. H., and Sunderland, R. J. (2008). Teacher turnover: Examining exit attrition, teaching area, transfer, and school migration. *Exceptional Children*, 75(1), 7–31.
- Boruch, R., Merlino, J., and Porter, A. (2013). Ambient positional instability: What we planned, what we learned in the first quarter, and a request. Presented at the National Forum on Education Statistics, March 6, 2013.
- Boruch, R. and Ruby, A. (2015) To Flop is Human: Inventing Better Scientific Approaches to Anticipating Failure. In R. Scott and S. Kosslyn (Eds) *Emerging Trends in the Social and Behavioral Sciences*. New York: John Wiley, pp 1-15. ISBN 978-118-90077-2 and http://repository.upenn.edu/gse_pubs/268.
- Bowdon, J. (2015). Teacher churn within schools: Magnitudes and implications for reform and inequality in student outcomes. Invited Presentation. Vanderbilt University, January 15 2015.
- Bowdon, J., and Boruch, R. (2014). Teacher Churn in Missouri's Biggest Cities, 2005–2014: A briefing. Philadelphia, PA: University of Pennsylvania, Graduate School of Education. Accessible at: http://repository.upenn.edu/gse/pubs_264.
- Cavanagh, S. (2011). Budget pressures churn workforce. *Education Week*
- Chao, Jessica., Park, Ji Euan, and Boruch, Robert. (2014). Ambient Positional Instability among Illinois Teachers: A briefing. Philadelphia PA: University of Pennsylvania, Graduate School of Education.
- Chan, Wendy (2016) Partial Identification of Treatment Effects: Application of Propensity Scoring to Generalizability. PhD Dissertation. Northwestern University.
- Frisone, Michael, Hooks, Thomas, Ye, Tianpeng, and Boruch, Robert (2016a). Ambient Positional Instability among Teachers in Arkansas Public Schools: Briefing to the Deputy Commissioner and Senior Administrators at the Arkansas Department of Education, Little Rock Arkansas, February 10 2016. Author: mfrisone@gse.upenn.edu.
- Frisone, M., Hooks, T., Ye, T., and Boruch, R. (2016b) Ambient Positional Instability among Teachers in Arkansas: A Report. Philadelphia PA: University of Pennsylvania, Graduate School of Education.
- Fullan, M. (2002). The implementation dip. *The Change Leader*, 59(8), 16–21.

- Fullan, M., & Hargreaves, A. (2012, June 6). Reviving teaching with “professional capital.” *Education Week*, 31(33), 30, 36
- Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Song, M., & Doolittle, F. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Graham, K. A. (2012, April 24). Phila. School District plan includes restructuring and school closings. *Philadelphia Inquirer*. Retrieved from <http://philly.com>
- Graham, K.A., Snyder, S., & Graham, T. (2011, Aug. 26). Ackerman and staff fell out of touch toward the end. *Philadelphia Inquirer*. Retrieved from <http://philly.com>
- Hanushek, E., Kain, J., & Rivkin, S. (2004). Disruption versus Tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics*, 88(9–10), 1721–1746. doi: 10.1016/S0047-2727(03)00063-X
- Hanson, T., Dietch, B., and Zheng, H. (2012). *Lessons in Character Impact Evaluation: Final report* (NCEE 2012-4004). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Heller, J. J. (2012). *Effects of Making Sense of SCIENCE™ professional development on the achievement of middle school students, including English language learners* (NCEE 2012-4002). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Marinelli, W. H. (2011). *The Middle School Teacher Turnover Project: A descriptive analysis of teacher turnover in New York City middle schools*. New York, NY: Research Alliance for New York City Schools.
- Mehana, M., and Reynolds, A. J. (2004). School Mobility and Achievement: A Meta-Analysis. *Children and Youth Services Review*, 26(1), 93–119. doi: 10.1016/j.childyouth.2003.11.004
- Merlino, F. J., and Musa, R. (2015). *Impact of teacher role instability on student performance in Pennsylvania*. Conshohocken, PA: 21PSTEM.
- Park, J. E., and Boruch, R. (2014). Ambient positional instability in randomized trials: A review of reports in the *Journal of Research on Educational Effectiveness*. Working Paper. Center for Research on Evaluation and Policy, Graduate School of Education, University of Pennsylvania
- Park, J. E., Boruch, R., Musa, R., and Chao, J. (2015) Record Matching in the Absence of Unique and Consistent Numerical Identifiers for Teachers in State’s Public Records: A Briefing.

- Porter, A., Merlino, J., Desimone, L., et al. (2012). Reports on the Cluster Randomized Trial on Enhancement of Science Curriculum for Middle School Math Students in Four Jurisdictions. AERA and others.
- Rampell, C. (2011, Nov. 8). More Churn in Job Market is Hopeful Sign. *The New York Times*. Retrieved from <http://economix.blogs.nytimes.com/>
- Ronfeldt, M., Loeb, S., and Wycoff, J. (2011). *How teacher turnover harms student achievement*. Washington DC: CALDER.
- Scull, Amanda Jane, and others (2015) Embedding Cognitive Science Principles into Science Instruction: Direct and Indirect Effects on Student Learning. Philadelphia: University of Pennsylvania, Graduate School of Education.
- Springer, M. G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D., ... Stecher, B. (2012). *Teacher pay for performance: Experimental evidence from the Project on Initiatives in Teaching*. Nashville, TN: National Center on Performance Incentives, Vanderbilt University. Retrieved from <http://www.performanceincentives.org>
- Stuart, Elizabeth, Cole, Stephen, Bradshaw, Catherine, and Leaf, Philip (2011) The Use of Propensity Scores to Assess Generalizability of Results from Randomized Trails. *Journal of the Royal Statistical Society*, 174 (Part 1, A), 369-386.
- Tipton, Elizabeth (2014) Stratified Cluster Sampling: A Sample Selection Strategy for Improved Generalizations from Experiments. *Evaluation Review* 37(2), 109-139.
- Tobias, R. (2012). Re: Piece for the New York Times/Email to Andrew Porter, April 26 4:40pm and Email to Robert Boruch April 30 10:04am.
- Useem, E., Offenber, R., and Farley, E. (2007). *Closing the teacher quality gap in Philadelphia: New hope and old hurdles*. Philadelphia, PA: Research for Action. Retrieved from <http://www.researchforaction.org>
- Vuchinich, S., Flay, R. R., Aber, L., and Bickman, L. (2012) Person Mobility in the Design and Analysis of Cluster Randomized Trials. *Prevention Science*, 33, 300-313. DOI 10.1007/s11121-011-0265-y
- Weijekumar, K., Hitchcock, J., Turner, H., Lei, P., and Peck, K. (2009). *A Multisite Cluster Randomized Trial of the Effects of Odyssey Math on the Math Achievement of Selected Grade 4 Students in the Mid-Atlantic region* (NCEE 2009-4068). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_20094068.pdf
- Yang, Rui, and others (2015) Applying Cognitive Science Principles to Middle School Science Curriculum: Comparing Two Approaches to Increasing Student Achievement. Philadelphia: University of Pennsylvania, Graduate School of Education.

Ye, Tianpeng, Frisone, Michael, Hooks, Thomas, and Boruch, Robert (2016a) New Jersey Public School Teachers' Retention, Instability, and Churn: 1996-97 to 2011-2012. Invited Presentation to the Deputy Commissioner and Senior Administrators, New Jersey Department of Education, Trenton, February 1 2016. Author: tianpeng@gse.upenn.edu.

Ye, Tianpeng, Frisone, Michael, Hooks, Thomas, and Boruch, Robert (2016b) New Jersey Public School Teachers' Retention, Instability, and Churn: 1996-97 to 2011-2012. Presentation at the Annual Meeting of the American Educational Research Association, Washington DC, April 2016. Author: tianpeng@gse.upenn.edu

Weathers, J., and Baker, J.Y. (2015, March). *Teacher positional instability in the Philadelphia public school system and the potential impact on interventions to improve student achievement*. Poster presented at the School District of Philadelphia Research, Policy, and Practice Conference. Philadelphia, PA.

Figures and Tables

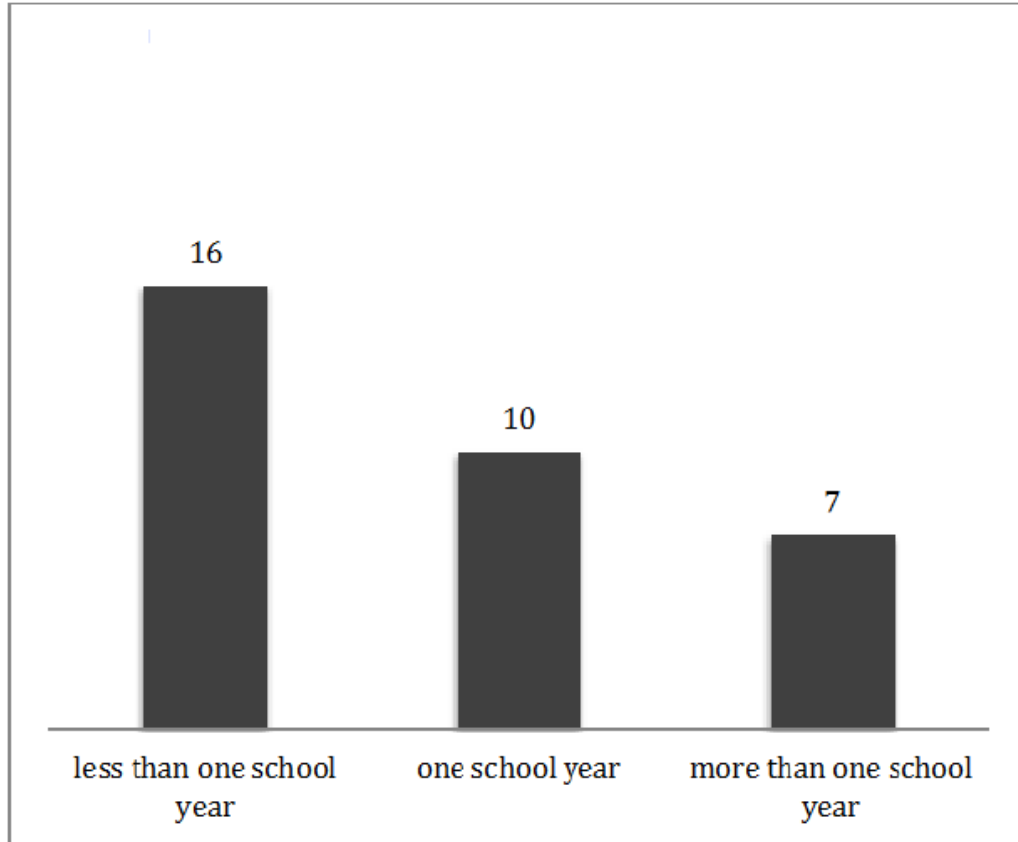


Figure 1. Duration of the API relevant interventions published in JREE between 2008-2013

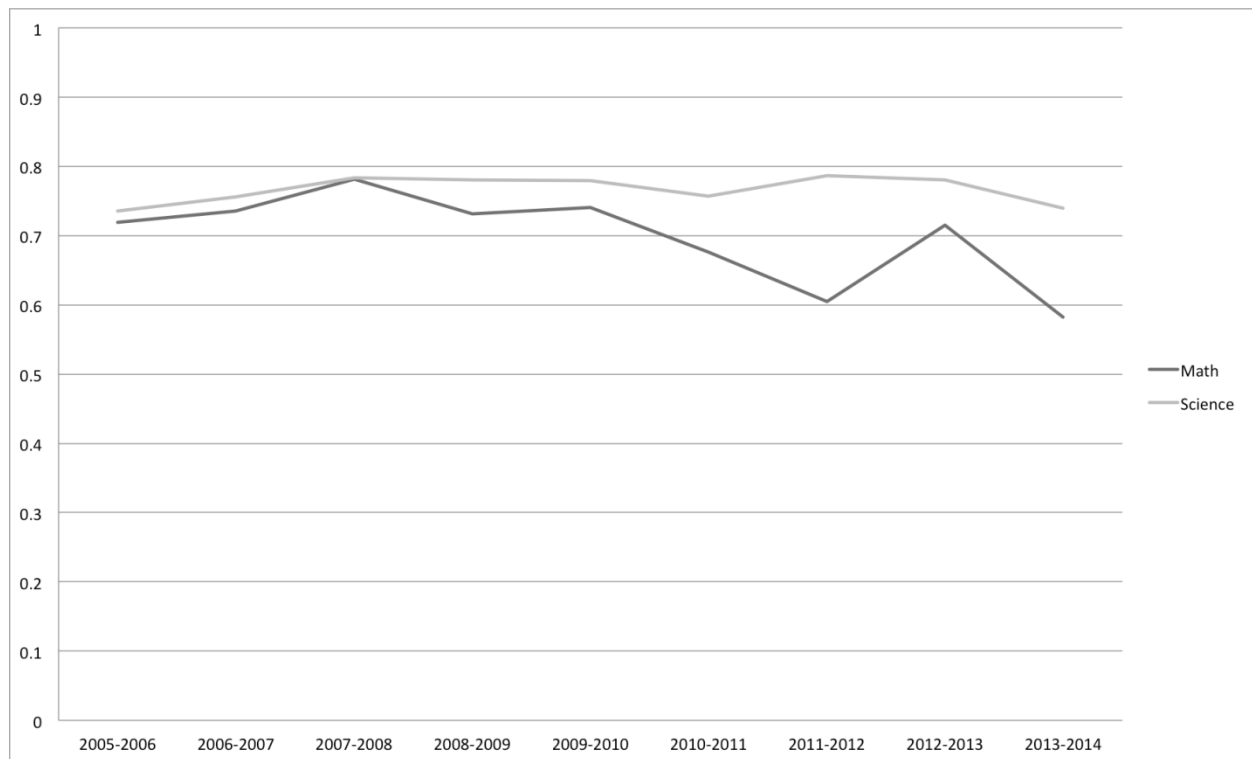


Figure 2 . Cohort Retention: Proportion of Middle and High School Teachers in the 5 Biggest Cities in Missouri Still Teaching the Same Subject Anywhere in the Missouri State System From One Year to the Next

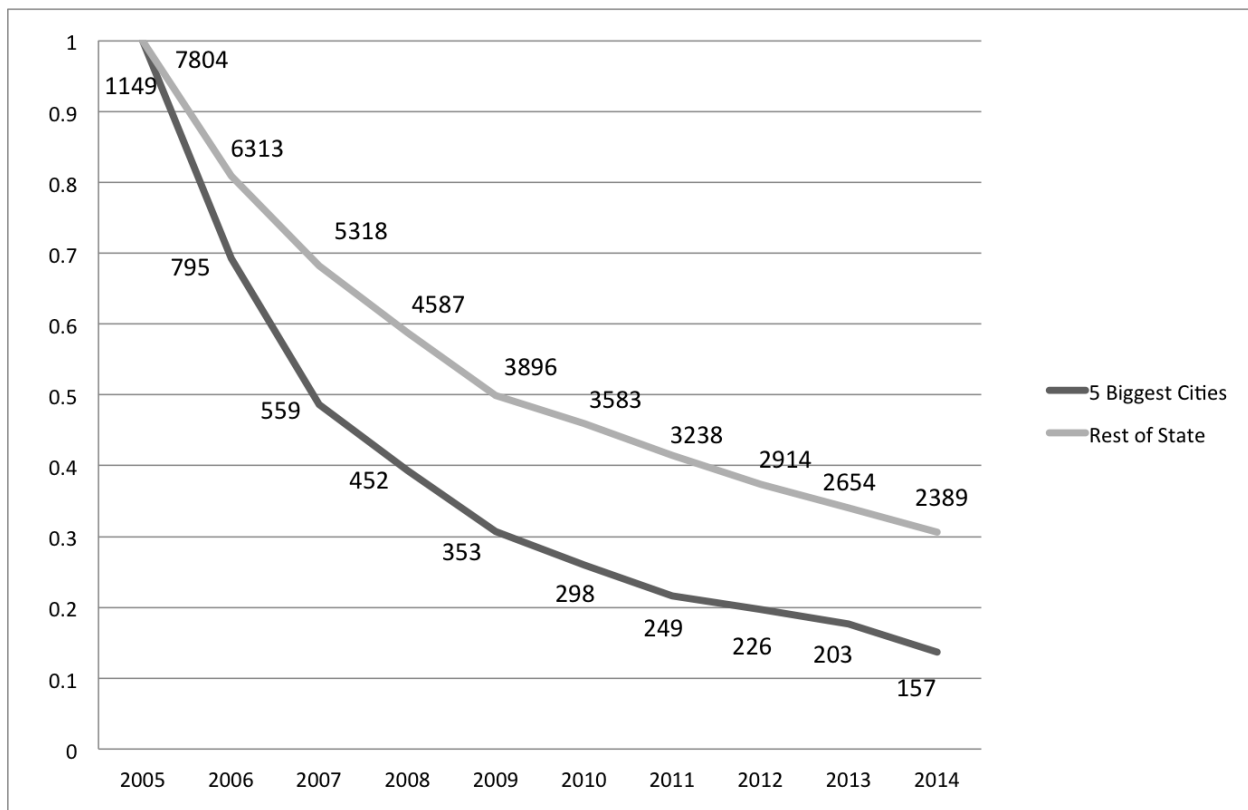


Figure 3. Cohort Retention: Proportion of STEM Middle and High School Teachers in Missouri in 2005 Still Teaching the Same Subject in the Same School Over a Ten Year Period

	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14
All Math and Science Teachers: In State	1465 100%	1209 83%	1043 71%	903 62%	753 51%	635 43%
All Math and Science Teachers: In Same District	1465 100%	1205 82%	1027 70%	882 60%	728 50%	608 42%
All Math and Science Teachers: In Same School	1465 100%	1102 75%	799 55%	657 45%	519 35%	417 28%
All Math and Science Teachers: In Same Subject(s) and School	1465 100%	1026 70%	728 50%	597 41%	469 32%	373 25%
Science Teachers: Same School	847 100%	602 71%	423 50%	348 41%	270 32%	220 26%
Math Teachers: Same School	892 100%	658 74%	476 53%	393 44%	312 35%	244 27%

Table 2. Cohort Retention: Teachers Remaining in the Same School from AY2007-2008 through AY2011-2012, for the State, Chicago School District, and the Five Next-largest School Districts*

AY	Statewide		Chicago		Five largest excl. Chicago	
2008	106954		19266		7027	
2008-2009	90741	84.8%	13416	69.6%	6072	86.4%
2008-2010	81813	76.5%	11551	60.0%	5460	77.7%
2008-2011	73281	68.5%	9904	51.4%	4776	68.0%
2008-2012	65649	61.4%	8733	45.3%	4197	59.7%

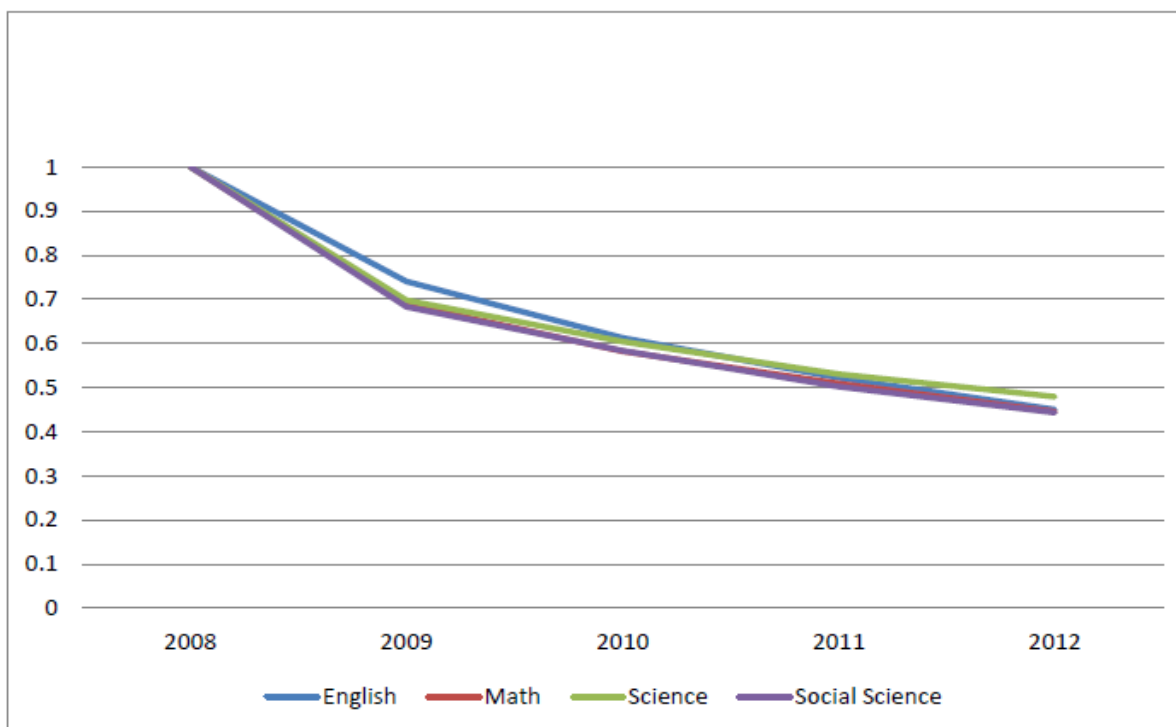


Figure 4. Proportion of Illinois teachers teaching the same subject area in the same school

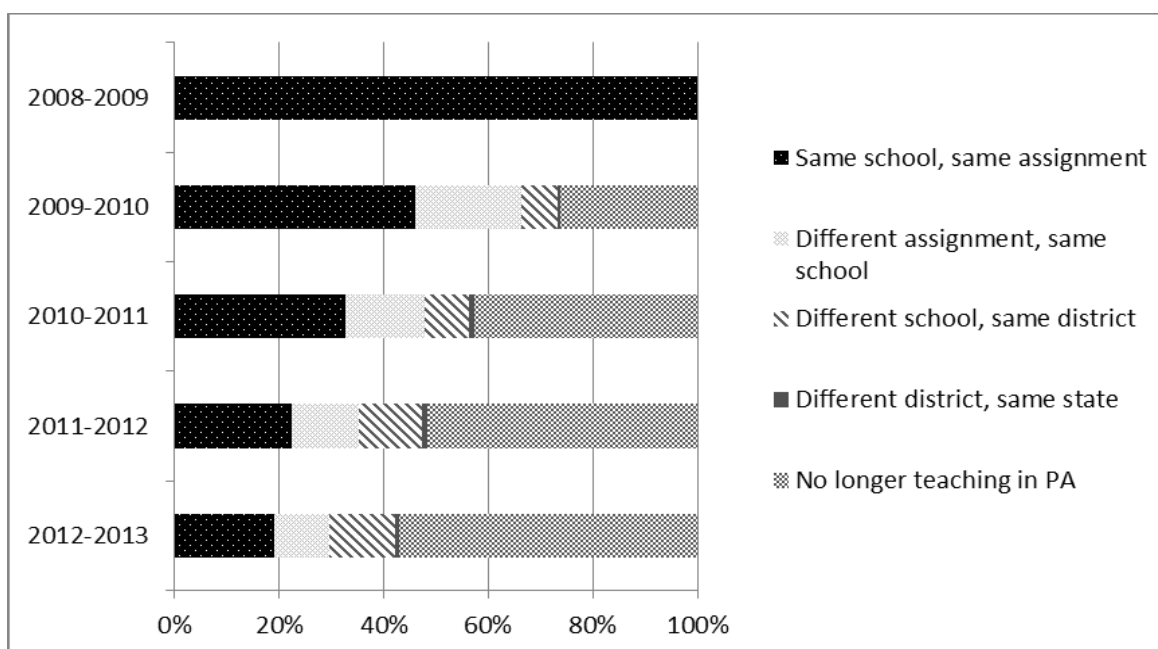


Figure 5. Math and Science Teachers in Public Non-Charter Secondary Schools in Phila: 2008-2009 Cohort Retention

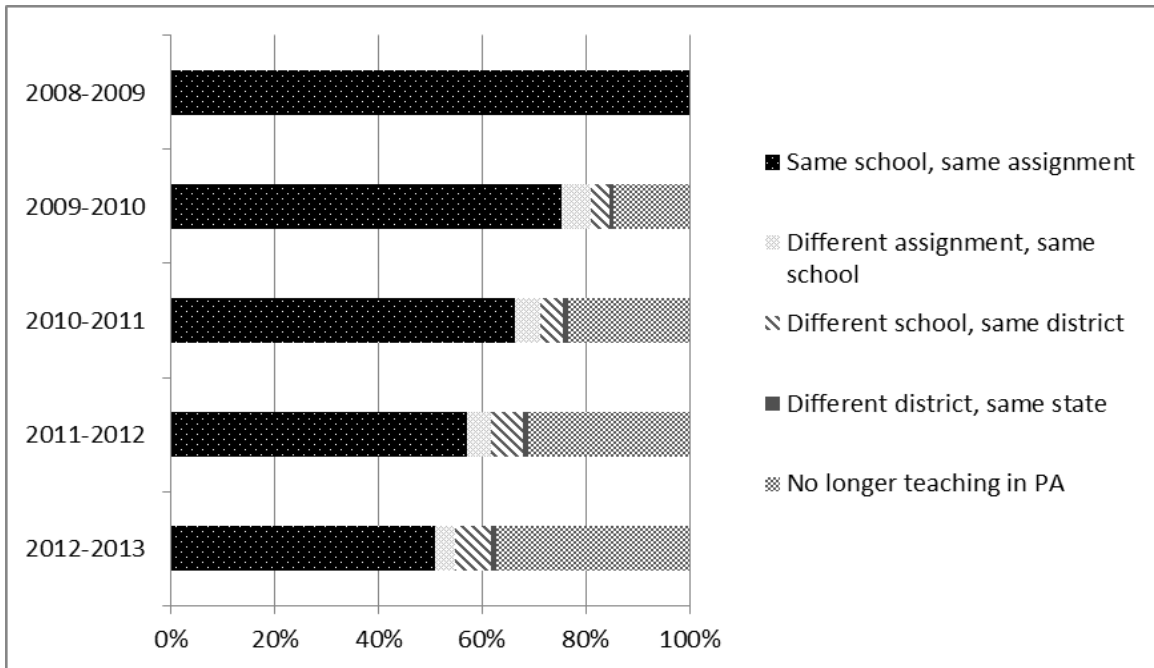


Figure 6. Math and Science Teachers in Public Non-Charter Secondary Schools in PA: 2008-2009 Cohort Retention

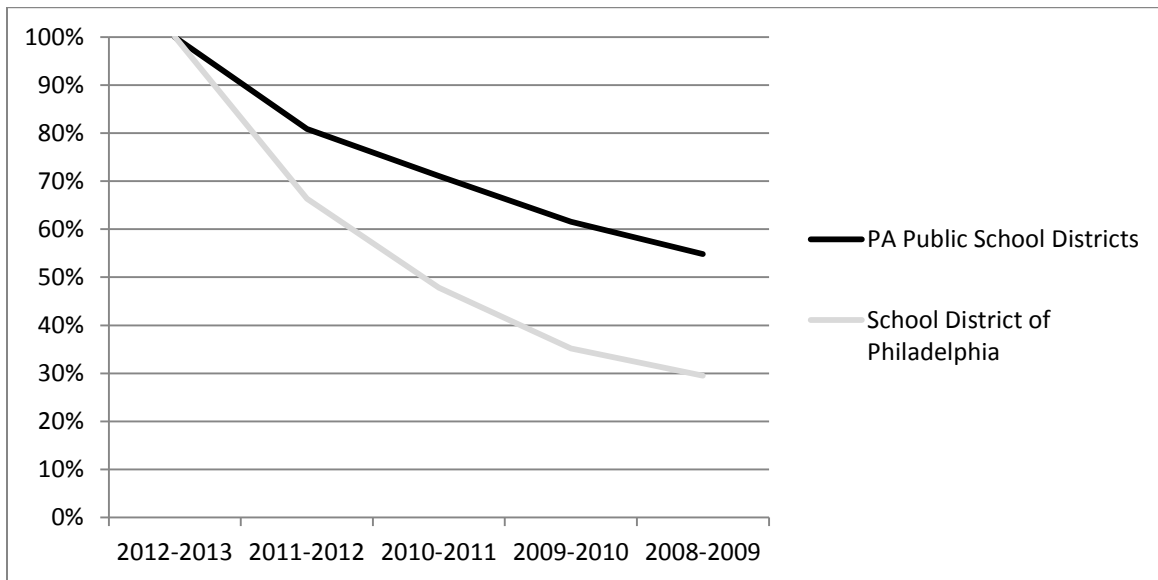


Figure 7. 2008-2009 Cohort Retention of PA Math and Science Secondary Teachers in the Same School

Table 3. Total number and proportion of all Arkansas high school teachers who remained teaching the same subject at the same school since 2010-11

		2010-11	2011-12	2012-13	2013-14	2014-15
Continues to Teach Math at the Same High School	Number of teachers retained from 2010-11	1742	1208	975	805	658
	Proportion of teachers retained from 2010-11	1.000	0.69	0.56	0.46	0.38
Continues to Teach Science at the Same High School	Number of teachers retained from 2010-11	1436	990	805	655	548
	Proportion of teachers retained from 2010-11	1.000	0.69	0.56	0.46	0.38
Continues to Teach English at the Same High School	Number of teachers retained from 2010-11	2198	1565	1253	1041	852
	Proportion of teachers retained from 2010-11	1.000	0.71	0.57	0.47	0.39
Continues to Teach Social Studies at the Same High School	Number of teachers retained from 2010-11	1492	1057	853	726	595
	Proportion of teachers retained from 2010-11	1.000	0.71	0.57	0.49	0.40
		2010-11 to 2011-12	2011-12 to 2012-13	2012-13 to 2013-14	2013-14 to 2014-15	
Unstable Teachers		7465	8157	8365	9064	
Number of Teachers		23312	23069	23059	23061	
Instability		0.32	0.35	0.36	0.39	

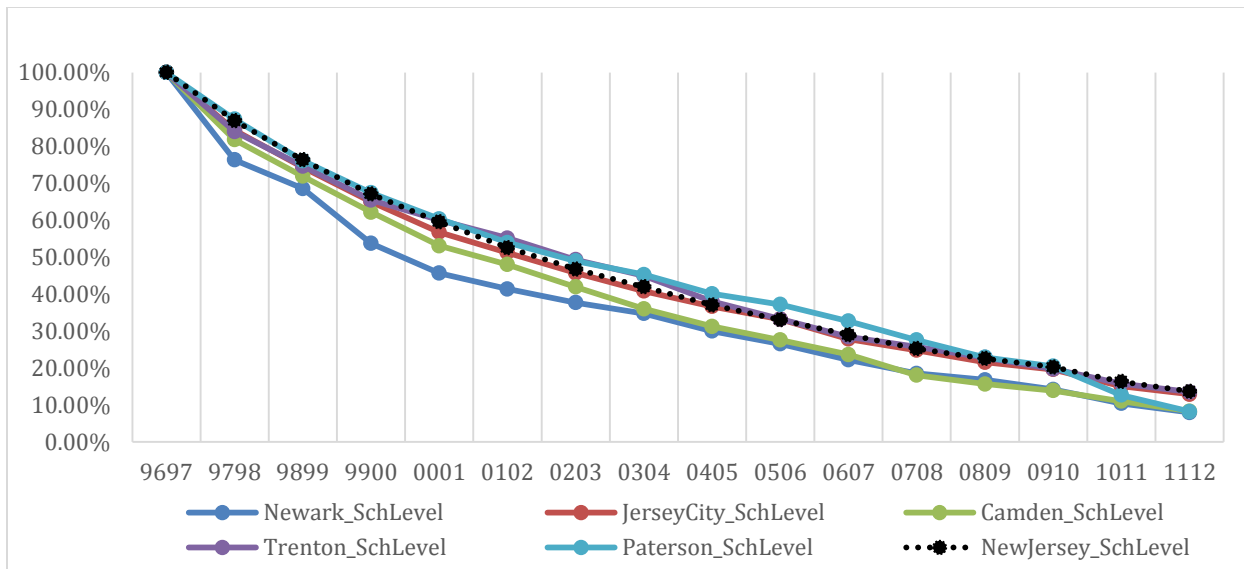


Figure 8. School-level retention by big cities

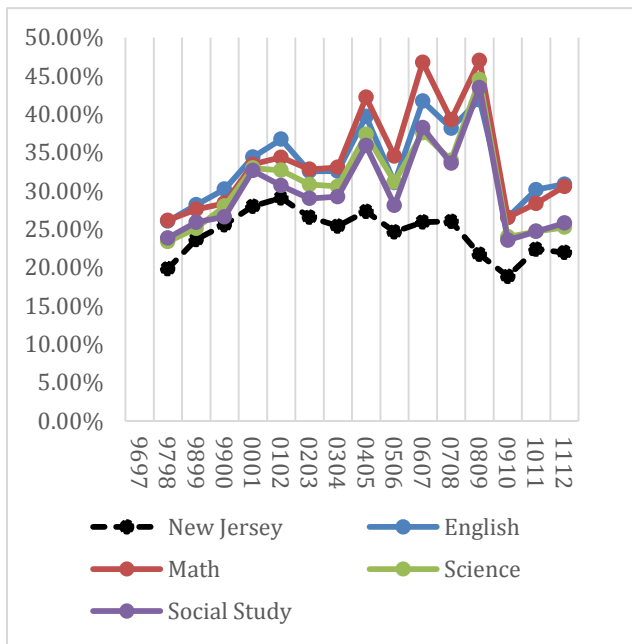


Figure 9. Churn by subjects